DOCUMENT RESUME

ED 367 635 SP 035 075

AUTHOR Seyfarth, John T.; And Others
TITLE Assessing Student Performance: Are Our Assumptions Valid?
PUB DATE Feb 94
NOTE 11p.; Paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education (Chicago, IL, February 16-19, 1994).
PUB TYPE Speeches/Conference Papers (150) -- Viewpoints (Opinion/Position Papers, Essays, etc.) (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Accountability; Achievement Tests; Cost Effectiveness; Educational Change; Elementary Secondary Education; *Evaluation Methods; Parent Attitudes; *Performance Tests; Standardized Tests; *Student Evaluation; Teacher Attitudes; *Test Validity; *Thinking Skills

ABSTRACT

Arguments for replacing standardized multiple choice tests with performance assessment that encourages teachers to devote more attention to higher order skills, and thus results in increased student achievement, are based on three assumptions: (1) the teaching profession, key decision makers, and parents will accept performance assessment measures as valid indicators of student achievement; (2) the use of performance assessment for accountability purposes will influence teachers to place more emphasis in their teaching on content that has significance for real-world tasks; and (3) the technical problems associated with developing performance assessments are solvable and the cost (in time and money) of this form of testing can be sustained. This paper examines evidence for the soundness of the three assumptions. The paper concludes that these assumptions may overlook critical facts about how professionals and parents are likely to respond to the introduction of performance assessment measures in schools, and suggests that a number of obstacles are likely to be encountered in the process of reforming assessment practices in schools. (Contains 12 references.) (JDD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 367 635

Assessing Student Performance: Are our Assumptions Valid?

John T. Seyfarth, Ed.D.
Professor

Diane J. Simon, Ph.D.
Associate Professor

Jeanne Schlesinger
Graduate Student

School of Education
Virginia Commonwealth University
Richmond, VA 23284-2020

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. T. Seyfarth

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Presented at the annual meeting of the
American Association of Colleges of Teacher Education
Chicago

February 1994

SP035075

Assessing Student Performance:
Are our Assumptions Valid?

Some people believe that replacing standardized multiple choice tests with performance assessment in schools will encourage teachers to devote more attention to teaching higher order skills and result in increased achievement by students. Their arguments are based on acceptance of three related assumptions about performance assessment and the factors that influence teachers' decisions regarding how and what they teach. The assumptions are:

1. The profession, key decision makers, and parents will accept performance assessment measures as valid indicators of student achievement

2. The use of performance assessment for accountability purposes will influence teachers to place more emphasis in their teaching on content that has significance for real-world tasks

3. The technical problems associated with developing performance assessments are solvable and the cost (in time and money) of this form of testing can be sustained

Whatever the inadequacies of standardized tests, that alone is not a sufficient justification for abandoning them unless it can be shown that the proposed replacements will solve the problems associated with existing tests. That will depend upon whether performance assessments can be successfully implemented in schools, which in turn will depend upon whether the three assumptions are valid. The purpose of this paper is to examine evidence for the soundness of the three assumptions.

Assumption 1

Educators who believe that multiple choice standardized tests are an unsatisfactory means of measuring student learning advance three arguments to support their position. These are, first, that standardized tests emphasize lower level recall and comprehension tasks and neglect higher order thinking skills such as problem solving and evaluation of information; second, that pressure from administrators and others to prepare students for the tests corrupts the educational process by causing teachers to teach to the test and neglect material not on the tests; and, third, that test content does not match the written or taught curriculum.

The first argument is rejected by the supporters of standardized tests, who argue that these instruments can and do assess students' ability to thinking critically ("Groups call for," 1990). The second and third claims contradict one another and cannot both be true. Either teachers adjust instruction to teach material on which students are tested, in which case the taught and tested curricula should conform, or they ignore test content and

teach other material, in which case there is at best only a poor fit between the two.

The question is whether teachers and parents will accept performance assessment measures as valid indicators of what students have learned and of how well schools are performing. There are reasons to believe that these groups may not readily embrace these measures.

Clearly the public wants access to data that show how well (or how poorly) schools are performing, and they believe that standardized multiple-choice tests serve that purpose. Few people without specialized training understand the technical objections raised by educators about multiple-choice standardized tests and are not particularly concerned about such issues as long as the tests appear to be valid measures of student achievement.

Performance assessment promotes the meaningful application of learned content, but previous efforts to emphasize application and deemphasize mastery of process skills have not always been well received. An example is the "new math" of the 1960s, when the mathematics curriculum was redesigned to focus on improving students' understanding of underlying concepts. Although some experts argue that the "new math" was never widely implemented in classrooms (Campbell & Fey, 1988), it was blamed in the public mind for students' poor mastery of computational skills. The experience demonstrated that professional educators and the public are not always in agreement regarding which instructional objectives should receive priority.

Clune (1993) cautioned that achieving consensus on the curriculum poses enormous social and political problems. The same might be said of testing. A research project sponsored by the Metropolitan Educational Research Consortium at Virginia Commonwealth University is investigating the experiences of school districts in the United States and Canada that have attempted to implement performance assessment. At least four districts which have adopted performance assessment have encountered opposition to the use of these tests from the public.

Parents who opposed the tests in those districts expressed concern that the introduction of performance assessment measures in schools would lead to a watering down of standards or to abandoning of traditional content. Some opposed performance assessment because of a genuine difference of opinion with educators regarding what the schools should teach. Many of these parents believe that teaching critical thinking skills is unnecessary and that testing students for recall and comprehension is acceptable.

Interestingly, most parents who oppose performance assessment do not appear to be concerned about the possible perversion of test scores in districts in which teachers are pressured to teach

specific test content as a way of raising students' scores, although that is one of the issues that has energized the effort to replace multiple-choice standardized tests. There is no evidence to suggest that performance assessments are more susceptible to misuse than standardized tests, and some people believe they are less subject to such abuse. However, a cynical public may accuse educators who favor performance assessments of preferring them because they expect them to make schools look better.

It seems likely therefore that educators can expect opposition from at least some parents and community members to proposals to substitute performance assessment instruments for the standardized tests now in use. Two of the districts surveyed as part of the MERC study also reported that they encountered resistance to performance assessment from a small number of teachers. In general, teachers are not strong supporters of mandated testing programs because they believe that testing takes time away from instruction and yields little information of value. Since performance assessments generally require more time to administer and score than the standardized tests now in use, opposition from teachers can be expected.

Most teachers, if asked, would probably agree that the knowledge and skills tested by well-designed performance assessment tasks are better measures of important learning outcomes than the multiple-choice items found on most standardized tests. It is not clear, however, whether teachers will change their instructional methods and content in response to a change in test format. That question is taken up in the next section.

Assumption 2

Grant Wiggins, a leader in the performance assessment movement, said that the belief that introducing a new type of test will induce teachers to change what or how they teach is "a hunch." "We'll see if it works out that way," he is quoted as saying (Rothman, 1989, p. 21).

Many educators believe that mandated multiple-choice standardized tests corrupt teaching by encouraging teachers to overemphasize unimportant test content while neglecting more important outcomes (Darling-Hammond, 1991). This charge is credible only if teachers' instructional decisions are influenced by the tests, and the evidence on that question is mixed.

Herman, Dreyfus and Golan (1990) cited three studies in which researchers reported that standardized tests had little effect on what teachers taught and an equal number that reached the opposite conclusion. In a study carried out by Herman and Golan (1993), the authors reported that teachers experienced strong pressure from district administrators and the media to improve their students' test scores. They added that teachers also reported a moderate

amount of such pressure from principals, other school administrators, other teachers, parents, and the community.

Moore (1992) found that elementary teachers in a Midwestern district that was ordered by a federal court to use the Iowa Test of Basic Skills to measure the effectiveness of the desegration effort based important decisions about what to teach on the content of that test. Majorities ranging from 70 to 100 percent of teachers in grades 3, 4 and 5 (N=79) said that they revised the curriculum scope and sequence, added lessons or units, increased the emphasis and amount of time devoted to material covered by the required tests, and eliminated certain topics in order to spend more time teaching content upon which students were to be tested. Eighty-seven percent or more of the teachers at all three grade levels indicated that they used test results to assess their teaching effectiveness (Moore, 1992).

Even allowing for social desirability bias, Moore's findings constitute impressive evidence of teachers' willingness to make adjustments in the way they allocate classtime, decide what to teach, and judge their own effectiveness in response to clear guidance from individuals in positions of authority.

However, court orders may have a more marked effect on teacher behavior than directives from a school official or publication of test scores in local newspapers, and Moore's (1992) findings reveal little about how teachers respond to those pressures. Teachers may well be more inclined to incorporate the instructional content from tests into their instruction when the decision to use the tests is made by a federal judge than when it comes from a school administrator, and endorsement of standardized tests by a federal court may lend them a legitimacy in teachers' minds they would otherwise not have.

It is not clear whether teachers are more subject to influence from pressure exerted by those in positions of authority or from their own beliefs about what knowledge is of most value. The issue is important because arguments for performance assessment generally assume that teachers ascribe greater inherent value to authentic tasks will therefore voluntarily devote more time and effort to preparing students for performance-type tests.

What can we learn from previous research? There is abundant evidence that teachers are influenced in what they teach by a number of external factors but less evidence that they are guided by strongly-held personal beliefs. In a survey study of teachers' attitudes about the effects of testing, Soltz (1992) concluded that elementary teachers "administered mandated standardized tests in ways largely uninfluenced by their personal feelings--negative or positive..." (p. 11).

Ongoing informal research with classes taught by one of the authors of this paper have produced a similar finding. Teachers in these classes have been surveyed for several years to determine whether they would be willing to teach a particular topic if it appeared in a textbook or they were asked to teach it, for example, by parents or the principal. The results suggest that teachers are surprisingly accommodating in response to requests to add new material to the curriculum.

Most of the respondents have said they would be willing to add the new content, even though they were aware that doing so meant they would have to delete other material. Their responses showed they are most strongly influenced by principals' expressed preferences about what to teach. These informal observations confirmed results of earlier studies that found that teachers readily acquiesced to pressure to add new content to the curriculum (Floden, Porter, Schmidt, Freedman, and Schwille, 1980).

On the other side, there is evidence that under certain circumstances teachers are prepared to resist administrative efforts to persuade them to emphasize test content in their instruction when they are not convinced of its value. Zancanella (1992) reported that principals' influence on instructional decisions of high school literature teachers were mediated both by teachers' attitudes about the tests and by their influence with colleagues.

The researcher concluded that teachers who disagreed with a principal's recommendation to prepare students for a mandated test and who had sufficient power with colleagues to feel comfortable in doing so successfully resisted the principal's entreaties, whereas those who agreed with the principal or perceived themselves as lacking the power to be able to resist went along with the principal.

One other factor that appears to affect how teachers respond to administrative appeals that has received relatively little attention in the research literature is teachers' attitudes about the ethical issues involved. Monsaas and Engelhard (1991) suggested that teachers' willingness to change instructional practices in order to prepare students for standardized tests were influenced by their attitudes about cheating and their perceptions of acceptable behavior. In a study involving 186 teachers, the authors found that teachers' attitudes about what constitutes cheating were better predictors of their responses to administrative requests than was the amount of administrative pressure they experienced.

In summary, findings from the few studies reviewed here suggest that the belief that introducing new tests in schools will result in changes in instruction is based on an oversimplified view of reality. Teachers' beliefs about the importance of the content

of mandated tests is only one of several factors that influence their decisions about whether to emphasize test content in their instruction.

Among the factors that determine how teachers will respond to mandated tests are their personal beliefs about the value of what is measured by the tests and their attitudes about the ethical implications of teaching to the test. Without more evidence on those issues, the assumption that the introduction of performance assessments in schools will produce desirable instructional change is tenuous. One conclusion that seems warranted, however, is that the introduction of performance assessment tests is most likely to lead to changes in instruction when care is taken in designing the tests to see that the content matches the curriculum and to ensure that the test results have value for teachers.

Assumption 3

Technical problems that are likely to be encountered in developing and implementing performance assessments in schools are of two types--those involving development and administration of the test instruments. Standardized multiple-choice tests have two appealing features that account for their continued popularity in spite of concerns about lack of content validity. The tests are high in reliability and low in both monetary and time costs. By their very nature, performance assessment instruments are lower in reliability and higher in cost than tests currently in use.

Concerns about reliability of performance assessment tests center around scoring consistency. Open-ended assessment tasks are usually scored by teachers, and maintaining consistency in scoring requires training scorers and providing for frequent reliability checks during the grading process. All of this increases the amount of time required to administer the tests and raises costs.

Some performance assessment measures are more difficult to administer than multiple-choice tests. Science assessments, for example, may require students to collect and interpret data, manipulate equipment or analyze substances. Teachers who administer the tests must set up the materials and equipment in advance of the test and remove and repackage them when testing is completed. The number of students who can be tested at one time when equipment is used is smaller than with pencil-and-paper tests, which means that more time must be set aside for administering the tests.

Writing assessments are designed so that students follow a model when completing a writing performance task, starting with a first draft which they then revise in subsequent sessions. In Arizona's statewide writing assessment, students write a draft one day and revise and edit it the following day using a checklist that is provided (Mitchell, 1992). While this arrangement allows for

contextual validity, it takes much more time than conventional tests. Administering the test over a two-day period also increases reliability concerns since students may use the time between writing and editing to collect additional information or to locate published sources from which they can borrow ideas and language.

None of these problems seems to be insurmountable. American schools probably administer more tests than necessary, especially in view of the fact that in many schools the results receive relatively little attention from teachers, and administering fewer tests but making better use of the information gleaned from them makes sense. However, despite the appeal of efficient and low-cost testing tools, educators, parents and taxpayers must face the fact that obtaining valid and reliable information about student achievement will involve greater costs in time and money than we have heretofore been willing to expend.

Conclusion

This paper has examined three assumptions about performance assessment and presented reasons to suggest that these assumptions may overlook critical facts about how professionals and parents are likely to respond to the introduction of performance assessment measures in schools. The paper has presented evidence suggesting that a number of obstacles are likely to be encountered in the process of reforming assessment practices in schools. Performance assessment may well prove to be a superior process of measuring student and school performance than current methods, but its potential will not be realized unless it is adopted and used. Rather than accept the assumptions without question, educators need now to initiate work to investigate more fully the conditions under which the assumptions are likely to prove to be true.

References

Campbell, P. & Fey, J. (1988). New goals for school mathematics. In R. Brandt (Ed.), Content of the curriculum (pp. 53-73). Alexandria, VA: Association for Supervision and Curriculum Development.

Clune, W. (1993). The best path to systemic educational policy: Standard/centralized or differentiated/decentralized? Educational Evaluation and Policy Analysis, 15, 233-254.

Darling-Hammond, L. (1991). The implications of testing policy for quality and equality. Phi Delta Kappan, 73, 220-225.

Floden, R., Porter, A., Schmidt, W., Freeman, D., & Schwille, J. (1980). Responses to curriculum pressures: A policy-capturing study of teacher decisions about content. East Lansing, MI: Michigan State University, Institute for Research on Teaching. (ERIC Document Reproduction Service No. ED 190526).

Groups call for phase-out of standardized tests. (1990, February 7). Report on Education Research, p. 5.

Herman, J., Dreyfus, J., and Golan, S. (1990). The effects of testing on teaching and learning. Los Angeles, CA: National Center for Research on Evaluation, Standards and Student Testing. (ERIC Document Reproduction Service No. ED 352382).

Herman, J. and Golan, S. (1993, Winter). The effects of standardized testing on teaching and schools. Educational Measurement: Issues and Practice, 20-25, 41.

Mitchell, R. (1992). Testing for learning: How new approaches to evaluation can improve American schools. New York: Free Press.

Monsaas, J., and Engelhard, G. (1991). Attitudes toward testing practices as cheating and teachers' testing practices. (Paper presented at the annual meeting of the American Educational Research Association, Chicago). (ERIC Document Reproduction Service No. ED 338643).

Moore, W. (1992). Testing perceptions, practices, and malpractice: The impact on teachers of court-ordered achievement testing in a desegregation setting. (ERIC Document Reproduction Service No. ED 344906).

Rothman, R. (1989, September 13). In Connecticut, moving past pencil and paper. Education Week, 1, 21.

Zancanella, D. (1992). The influence of state-mandated testing on teachers of literature. Educational Evaluation and Policy Analysis, 14, 283-295.